First I was trying to figure out which Python libraries I was going to need in order to achieve this project, however, I noticed that trying to figure that out before actually starting to code the application, was not very realistic, instead I realized it was better to just import them as I started to find a need for them.

Processing the csv file that was already provided to me was the easiest to handle, given that the file was already gathered. The file generated from Twitter's API was very difficult to obtain due to my inability to obtain a Twitter developer account. I was unable to follow through their process, instead I was provided with a file with the JSON response from the API we should have run which I was later able to parse through and create a data frame from it.

Assessing the data was my most difficult task, there was a lot of data to look through and understand how it could be better. I spotted easy quality issues such as columns having inappropriate data types and common columns between tables that could be merged together. Nonetheless, there were more complex quality issues to address, such as the issues with the image predictions, where the AI had made predictions in regards to the images in which included items that were not dog breeds, there were many predictions that did not make sense, and trying to correct that would be a rigorous task that I decided to avoid and focus my efforts in improving other quality issues.

Finally cleaning the data was relatively easy, given that while I was assessing, I was already thinking about how I would address these concerns. I changed data types, I merged tables, I dropped columns and constantly check my data frames to make sure my changes were applied. I found myself going back to the "Assess stage" after my cleaning was done, just to verify my changes were actually effective. Once everything was cleaned up, I generated a visual and 3 insights from this data. I feel the project may have been more interesting for my taste if it was related to economics, or medical data, something a little closer to reality that I could use to better understand the world that we live in.